Velasco, J.A., Amaris, H., Alonso, M. y Casas, M. (2019). Energy losses estimation tool for Low Voltage Smart grids. In *CIRED Conference Proceedings - 25th International Conference on Electricity Distribution.*

ENERGY LOSSES ESTIMATION TOOL FOR LOW VOLTAGE SMART GRIDS

José Angel VELASCO
Universidad Carlos III de Madrid - Spain
joseangel.velasco@uc3m.es

Hortensia AMARÍS
Universidad Carlos III de Madrid - Spain
hortensia.amaris@uc3m.es

Mónica ALONSO
Universidad Carlos III de Madrid - Spain
monica.alonso@uc3m.es

Marta CASAS Naturgy
Spain
mcasas@gasnatural.com

ABSTRACT

The so-called 20-20-20 targets committed to by the European Union drives the need for a more efficient distribution network. The energy efficiency improvement required involves a 20% reduction of the energy consumption compared to the 1990s. For such cutback, Distribution Systems Operators are encouraged to develop the best strategies to identify and reduce power losses in their networks. This task becomes challenging in Low Voltage Distribution networks due to diversity in the feeder's topology configuration, load distribution and the presence of renewable-based distributed generation. In this paper, a clustering-based methodology is proposed as an energy losses tool to support the energy efficiency decision-making process. A feeder's clustering process using the K-means algorithm is carried out upon a customised network characteristics set that was previously reduced to two coordinates by applying Principal Component Analysis. The relationship between power losses and the net energy imported under the different scenarios is obtained for each feeder class identified. The data and network used in this process correspond to the roll-out deployed at the Spanish Smart Grid Demonstration Project (OSIRIS).

INTRODUCTION

Due to environmental concerns, in 2007, the Europe Union (EU) committed to reaching the so-called 20-20-20, which consists of reducing 20% the Greenhouse Emissions (GHE) compared to the 1990s, increasing the energy produced from Renewable Energy Sources (RES) by 20% and reducing energy consumption by 20%. In this context, the electricity distribution sector plays a fundamental role, especially in Low Voltage (LV) distribution networks, where it has produced a higher rate of power losses compared to transport networks. Therefore, to achieve these EU objectives, Distributed System Operators (DSOs) are encouraged to implement optimal strategies to locate the inefficient power loss sources in their networks, and then to apply the required actions to minimise those, thereby achieving the highest rate of energy efficiency. However, LV distribution networks include multiple network topologies (e.g., overhead and underground), lateral branches (mostly radial structure), multiple customer types, and Distributed Generation (DG) units, such as Photovoltaic (PV) facilities connected to the network. Moreover, despite the deployment of smart metering systems in recent years, which have changed the traditional LV network into the well-known Smart Grids [2], there still exists a great level of uncertainty related to the network information at a topology configuration level. In this situation, the evaluation of energy efficiency of a network becomes challenging due to both the difficulty of modelling every single feeder (due to this uncertainty) and the uncertainty of the load demand (which produces the power losses). In the scientific literature, a large number of works have addressed losses calculation in distribution networks. Table 1 shows a review of the most recent research papers on this topic.

Table 1.: Losses calculation methods in the scientific literature

References	Methodology	Limitations
[3]-[4]	Loss and Load factors	Requires accurate demand data and it ignores network topology
[5]	Top down /bottom up approach	Requires accurate demand data
[6]-[8]	Load flow analysis and Regression Analysis	Deep knowledge of the network topology is required
[9]	Clustering techniques	High computational complexity

From the papers cited, it is clear that a detailed knowledge of the network is required to perform the losses calculation. Given that this information is sometimes limited or difficult to access, much research effort has been aimed at overcoming this limitation. In this paper, a clustering approach is proposed as an energy loss estimation tool for Smart Grids. This approach identifies those groups of feeders that share statistical similarities, which can be used for obtaining the relationship between the energy imported and the energy losses on a daily basis. The output of the energy losses estimation tool is synthesised using the relationship between the energy imported by the LV feeders from the network and the energy losses that are produced in that network as a consequence of the energy supply of the customers, as well as the power injections of the PV units distributed. Nonetheless, this relationship depends on many factors, such a topology configuration of the feeder (presence of lateral branches or far away laterals), type of customers (customers with different power contracted), and their distribution along the feeder (customers with high power contracted close to the substation of at the end of the feeder), presence of distributed resources (such PV units) and their location along the feeder, and unbalanced loading, among others. This heterogeneity makes the estimation of power losses arduous due to the need to model every feeder. On the other hand, losses could be subtracted from the readings of all of the smart meters of the customers and the supervisor meter located in the secondary substation. However, this raises some technical challenges due to the existence of non-telemetered customers, who provided consumption records on a monthly basis, errors in the readings and billings and the presence of illegal connections from which the DSO is unaware.

METHODOLOGY

The losses estimation tool proposed in this paper works as follow. First, a set of characteristic parameters of the feeders are proposed to reflect the different topologies and configurations. These characteristic parameters are collected from real LV networks. Second, by applying Principal Component Analysis (PCA), this set of parameters is reduced to two coordinates [11] retaining the variance of the data. Next, a k-means clustering process [12] is carried out to obtain the clusters or feeders. Then, for each cluster, a representative feeder is selected, and the relationship between the energy losses and the energy imparted to the feeder is obtained by running a Monte Carlo simulation under different load and generation scenarios.

Characteristic Parameters

The parameters proposed to capture the variability of the LV feeders are the following:

- **Laterals Weight LW_i**: The total sum of active power demand of the customers connected to a lateral branch weighted with the length of the lateral, in relation with the total power demand of all the customers connected. This is calculated as indicated in (1), where $p_{d,k,lb}$ is the active power demand by customer k connected to the lateral branch l_b, n_c is the number of customers connected to lateral, n_{lb} is the number of laterals of the feeder, l_{lb} is the length of the lateral branch, L_i is the length of the feeder, and N_c is the total number of customers connected to the feeder.

$$LW_i = \frac{\sum_{lb}^{n_{lb}} \sum_{k}^{n_c} p_{d,k,lb} \, (l_{lb}/L_i)}{\sum_{k}^{N_c} p_{d,k}} \qquad (1)$$

- **Laterals Location LL_i**: The average location of the laterals along the feeder. The distribution of the laterals has a great impact on the power losses due to the radial topology of the feeder. It is calculated as indicated in (2), where d_{lb} is the distance between the connection point of the lateral with the feeder and the head of the feeder (secondary substation)

$$LL_i = \frac{\sum_{lb}^{n_{lb}} d_{lb}}{L_i} \qquad (2)$$

- **Load Distribution LD_i**: The distribution of the power demand along the feeder**.** This parameter takes a value equal to "0" if the distribution of the load is uniform, "-1" if the distribution is decreasing and "1" if the distribution is increasing (Figure 1).

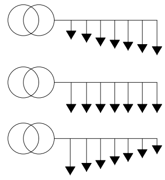

Figure 1.: Load distribution parameter

- **PV Generation Level PVL_i**: Measures the influence of the PV facilities connected to the feeder. This is calculated as the ratio of the power peak of each installation and the total power demand of the feeder, as indicated in (3), where N_{PV} is the number of PV facilities connected to the feeder and $p_{PV,k}$ is the peak power of the facility under standard conditions (STC).

$$PVL_i = \frac{\sum_{k}^{N_{PV}} p_{PV,k}}{\sum_{k}^{N_c} p_{d,k}} \qquad (3)$$

- **PV Generation Distribution PVD_i**: The distribution of the connection of the PV facilities along the feeder. This parameter takes value in the same way as LD_i.

- **Unbalanced Load Level $ULL_{i,p}$**: Measure the unbalance operation condition of the feeder. It is calculated as the load demand connected to the phase p in relation to the total demand of the feeder as it is indicated in (4), where $N_{c,p}$ is the customers connected to the phase p and p_d^p the power demanded by them.

$$ULL_{i,p} = \frac{\sum_{k}^{N_{c,p}} p_d^p}{\sum_{k}^{N_c} p_{d,k}} \qquad (4)$$

In the absence of large sources of reliable information about LV network topologies, in this paper, a heuristic topology builder algorithm has been developed to create sufficient samples of feeders to use in a clustering process. The algorithm design is based on the expertise acquired in the OSIRIS [13] research project and reproduces the characteristics of the distribution network topologies of a large distribution area of Madrid (Spain). The feeders are underground and consist of aluminium cables with a cross-section area of 240 mm^2. The distribution of the characteristic parameters of the feeders synthetically created is shown in Figure 2.

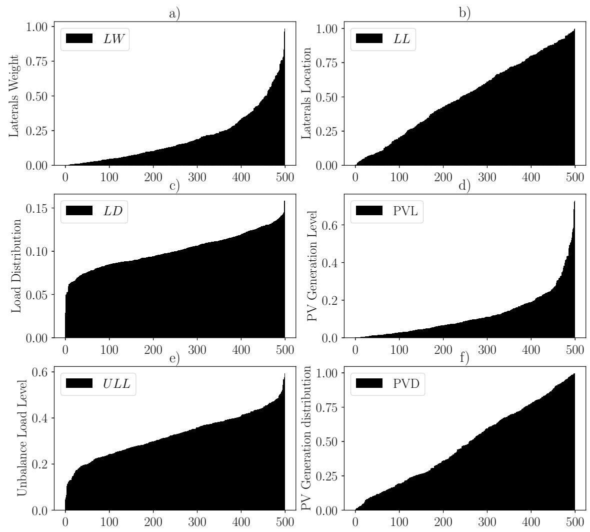

Figure 2.: Characteristic parameters of the synthetically generated feeders

Representation system

The above set of characteristic parameters is reduced to two coordinates (principal components) using an orthogonal linear transformation called PCA [11], widely used in exploratory data analysis for building predictive models. This transformation is done in the sense that the first coordinate has the largest variance (e.g., the first coordinate captures the maximum possible variability of the data) and the second coordinate has the highest variance subject to the constraint that it is orthogonal to the first component. Assuming that the characteristic parameters are arranged in a matrix M_{nxp} with n rows corresponding with observations (feeders created) and p columns corresponding with the six parameters, the transformation is defined by a p-dimensional weighted vector $\varphi_{(k)} = (\varphi_1, \dots, \varphi_p)_{(k)}$ that transforms each row $x_{(i)}$ of the matrix M_{nxp} into a new component $t_{k(i)}$, as indicated in (5), where l is the number of principal components, and in this case l =2.

$$t_{k(i)} = x_{(i)}\varphi_{(k)}, \forall i \in (1, \dots, n), \forall k \in (1, \dots, l) \qquad (5)$$

The first coordinate is obtained by maximising the variance, so the first weighted vector $\varphi_{(1)}$ corresponding with the first coordinate is calculated, as indicted in (6), and the first coordinate is obtained as indicated in (7).

$$\varphi_{(1)} = \arg max_{||\varphi||=1} \left\{ \frac{\varphi^T M^T M \varphi}{\varphi^T \varphi} \right\} \qquad (6)$$

$$X_{1(i)} = x_{(i)}\varphi_{(1)} \qquad (7)$$

To calculate the second coordinate, it is necessary to subtract the first component from the matrix M_{nxp} as indicated in (8), and then find the weighted vector that maximizes the variance of the new matrix $\widehat{M}_{nxp}$, as indicated in (9), and finally the second coordinate is calculated, as indicated in (10).

$$\widehat{M}_{nxp} = M_{nxp} - \sum_{s=1}^{k-1} M_{nxp} - \varphi_{(s)}\varphi_{(s)}{}^T \qquad (8)$$

$$\varphi_{(2)} = \arg max_{||\varphi||=1} \left\{ \frac{\varphi^T \widehat{M}^T \widehat{M} \varphi}{\varphi^T \varphi} \right\} \qquad (9)$$

$$Y_{1(i)} = x_{(i)}\varphi_{(2)} \qquad (10)$$

Clustering Process

Once the feeders are represented with two coordinates, a clustering procedure using K-means algorithm [12] is performed to obtain the groups of the feeders that have similar statistical properties (clusters). By this way the n observations obtained with the topology builder algorithm are partitioned in K sets. Mathematically this is to find the sets that minimise the pairwise squared distances of points in the same clusters. The process consists of three steps. Initially in the initialization step, K points from the total set of points are chosen randomly to be the initial centroids (centres of the cluster). In the next step (initial cluster assignment) the initial clusters are created with all the data points closest to the centroids. In this step Euclidean Distance (12) is used to measure that notion of distance. After that initial steps, the centroids are updating for the clusters defined. These last two steps are repeated until the position of the centroids converges.

$$d(p,q) = \sqrt{(X_{1,p} - X_{1,q})^2 + (Y_{1,p} - Y_{1,q})^2} \qquad (12)$$

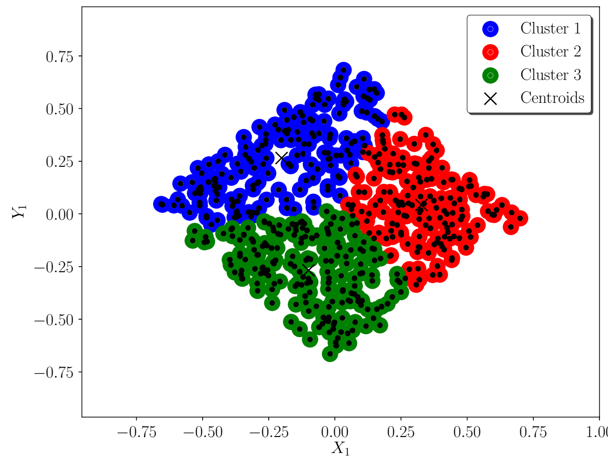

Figure 3.: Feeders clusters obtained applying K-means algorithm

The number of clusters is chosen by testing different values of K and then by evaluating the performance of the clustering done in terms of variability, which is defined as the sum of all Euclidean distances between the centroid and each data point. Finally, Figure 3 shows the representation of the synthetic feeders by means of the two principal coordinates explained above. Also, the results from the clustering process are indicated, where the clusters are indicated with colour circles and the centroids are indicated with crosses. It can be seen that three different clusters of feeders have been identified. The above graph could be generalized in the following classification map when applying new feeders (Figure 4).

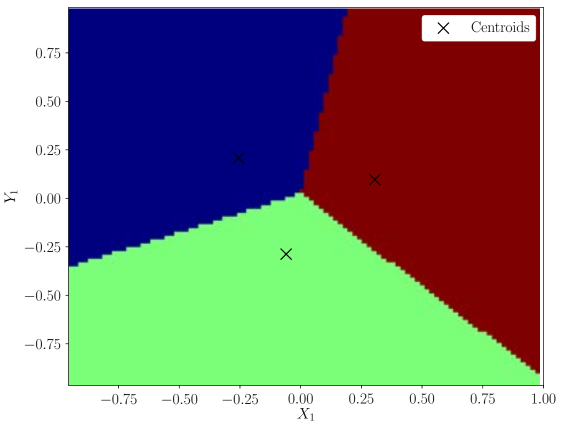

Figure 4.: Feeder Classification Map

Power losses calculation

After obtaining the different clusters of feeders, a representative feeder from one cluster is selected–being the closest to the centroid. Therefore, with these three representative feeders, a Monte Carlo simulation is carried out with different demand and PV generation scenarios. This simulation consists of setting up the load scenarios in each representative feeder and then solving the load flow problem and obtaining the losses in relation to the load demand. The load flow is solved by considering a balanced situation. Notice that this is not a loss of accuracy since the unbalanced conditions were considered in the classification process. Three scenarios are considered in a function of the power demand by the customers, which is referred to as the mean value and standard deviation (std). Meanwhile, PV presence has been kept uniform along the feeder. Scenario 2 corresponds with the mean value of load demand, while Scenarios 1 and 3 correspond with the load demand equal to the mean value plus a standard deviation and minus a standard deviation, respectively.

The results of this simulation are gathered and fitted by applying a polynomial regression model [14], as indicated in (13), where E_L is the energy losses in the network by joule effect, E_{imp} is the energy imported to the feeder from the MV network, and $\beta_0, \beta_1, \beta_2$ are the coefficients to be fitted.

$$E_L = \beta_0 + \beta_1 E_{imp} + \beta_2 E_{imp}{}^2 \qquad (12)$$

The energy imported is the independent variable and can be obtained from the available energy measurements that provide the smart meter located in the secondary substation. In Figure 5, we show the daily energy losses obtained in the Monte Carlo Simulation fitted with a polynomial regression curve referred to the daily energy demand expressed in per unit (considering 1 MW the reference for the energy imported). This shows a quadratic behaviour of the energy losses with respect to the energy imported.

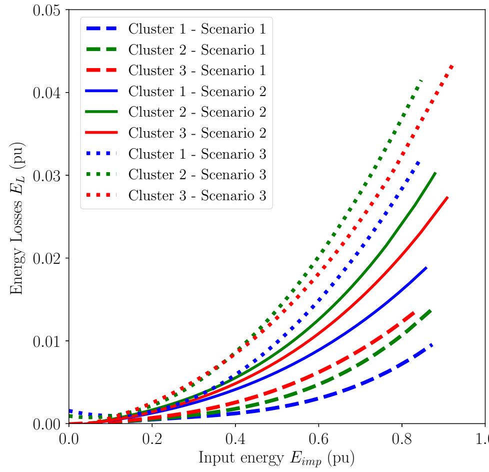

Figure 5.: Energy losses- input energy relationship for each cluster and scenario.

Therefore, with the above quadratic curves and the feeders classification map presented in Figure 4, it is possible to estimate the power losses of new feeders by knowing the energy imported and the cluster where it belongs.

CASE STUDY

Description

The proposed losses estimation tool has been applied in the Spanish demonstration project OSIRIS [12]. The distribution power network of the OSIRIS pilot is heterogeneous and involves different LV networks (rural, urban and semi-rural). The project scenario concerns a primary substation that supplies power to 31,000 residential and industrial customers in a region located in the south of Madrid (Spain) with a total contracted power of 155 MW distributed in 750 feeders having an accumulated length of 164 km.

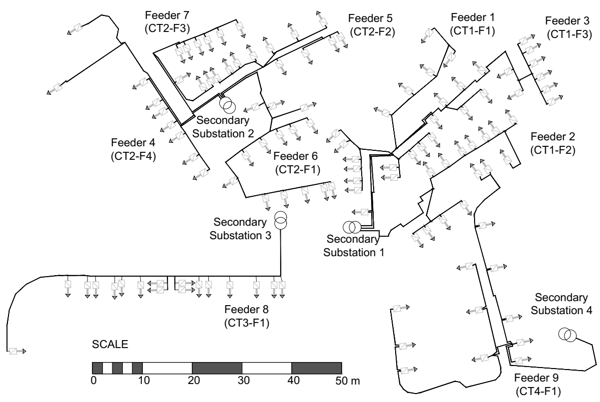

Figure 6.: OSIRIS network layout

Results

To illustrate the process of the proposed losses estimation tool, a set of seven representative feeders from the OSIRIS distribution networks were selected and are shown in Figure 6, where all customers have a contractual power of 15 kW, and the power factor is 0.9 (inductive). The nine feeders have been classified, and the power losses have been estimated using the quadratic expressions obtained, as shown in Figure 7, this results in an error in comparison with load flow of less than 5%.

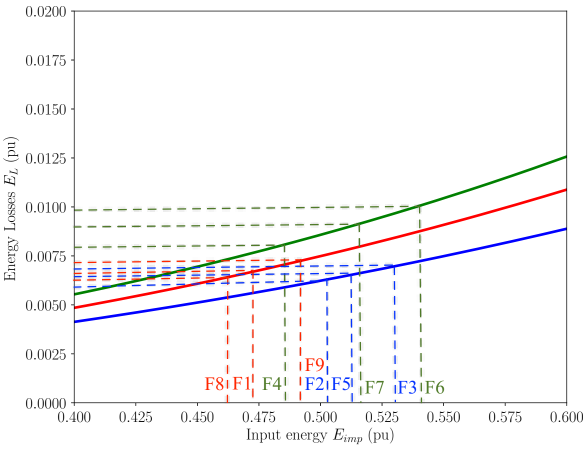

Figure 7.: Losses estimation of the OSIRIS feeders

CONCLUSIONS

A losses estimation tool for smart grids has been proposed. A set of representative parameters has been proposed to characterise the feeders that compound the smart grid. Then, a dimensionality reduction is carried out using PCA to represent the feeder with two coordinates. Next, a clustering process using the k-means algorithm has been conducted to obtain three different feeders clusters. With this, a feeder's classification map has been built. Consequently, the relationship between the energy losses that take place in a feeder and the energy imported as a result of the load demand and the PV generation have been found through a Monte Carlo simulation. In this simulation, the variability of the demand has been considered via scenarios. Finally, the usefulness of the proposed tool was validated with a set of representative feeders belonging to a demonstration smart grid project. The results demonstrate the convenience of the presented tool as an alternative to the classic load flow approach as the approach to obtain the power losses of a smart grid LV feeder.

Acknowledgments

This work has been partly funded by the Spanish Ministry of Economy and Competitiveness through the National Program for Research Aimed at the Challenges of Society under the project OSIRIS (RTC-2014-1556-3).

REFERENCES

[1]. Directive 2012/27/EU of the European Parliament and the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8//EC and 2006/32/EC.

[2]. G. López, J. I. Moreno, H. Amarís and F. Salazar, 2015, "Paving the road toward Smart Grids through large-scale advanced metering infrastructures," *Electric Power Systems Research*, vol. 120, pp. 194-205

[3]. L. M. O. Queiroz, M. A. Roselli, C. Cavellucci and C. Lyra, 2012, "Energy Losses Estimation in Power Distribution Systems," *IEEE Transactions on Power Systems*, vol. 27, no. 4, 1879-1887

[4]. X. Fu, H. Chen, R. Cai and P. Xuan, 2016, "Improved LSF method for loss estimation and its application in DG allocation," *IET Generation, Transmission & Distribution*, vol. 10, pp. 2512-2519.

[5]. K. A. Ibrahim, M. T. Au, C. K. Gan, and H. J. Tang, 2017, "System wide MV distribution network technical losses estimation based on reference feeder and energy flow model," *Int. J. Electr. Power Energy Syst*, vol. 93, pp. 440–450.

[6]. C. A. Dortolina and R. Nadira, 2005, "The loss that is unknown is no loss at all: a top-down/bottom-up approach for estimating distribution losses," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1119-1125.

[7]. E. Mckenna, I. Richardson, M. Thompson, 2012, "Smart meter Data: Balancing consumer privacy concerns with legitimate applications," *Energy Policy*, vol. 41, pp. 817-814.

[8]. A. J. Urquhart and M. Thomson, 2015, "Impacts of Demand Data Time Resolution on Estimates of Distribution System Energy Losses," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1483-1491.

[9]. G. Poursharif, A. Brint, M. Black, M. Marshall, 2018, "Using smart meters to estimate low-voltage losses," *IET Generation, Transmission & Distribution*, vol. 12, pp. 1206-1212.

[10]. S. Wang, P. Dong and Y. Tian, 2017, "A Novel Method for Statistical Line Loss Estimation for Distribution Feeders Based on Feeder Cluster and Modified XGBosst," *Energies*, vol. 10.

[11]. I. T. Jolliffe, 2002, *Principal Component Analysis*, Springer, Aberdeen, UK,

[12]. G. Gan, C. Ma and J. Wu, 2007, *Data Clustering: Theory, Algorithms and Applications*, SIAM.

[13]. OSIRIS Project www.proyecto-osiris.com

[14]. D.C. Montgomey, E. A. Peck and G. G. Vinning, 2012, *Introduction to linear regression analysis,* Wiley.